How to Grow an Astroinformatician

Petr Škoda

Astronomical Institute, Czech Academy of Sciences Ondřejov

Supported by COST Action TD-1403
BIG-SKY-EARTH
and
grant LD-15113 of the Czech Ministry of
Education, Youth and Sports

Presented at COST Conference joined with IAU Sympozium 325
on Astroinformatics
25th October 2016, Sorrento, Italy

CV and Experience

- Stellar astronomer
  - Interest in IT always
  - Theory (HII regions RRL, Relikt < COBE)
  - Astronomical Instrumentation
  - Big telescopes – control systems
  - CCD – detectors , processing of RAW data
  - Echelle spectroscopy
  - Virtual observatory, data archives
  - Astroinformatics, ML
  - Student's Bc, MS. Theses – Informaticians

CV and Experience

- Milestones:
  - 1994 Weaving the Astronomy Web in Strasbourg
  - 1995 ADASS (Tucson)
  - 1997 IAC Winter school – Astronomy in the Internet Age
  - 1998 Grant on building web archive
  - 2003 Virtual observatory (2003 ADASS – IVOA)
  - 2006 IAU GA– EURO-VO involvement + NVO School
  - 2007 workshop on publishing in VO
  - 2007 ADASS tutorial on ML (S. McConnel)
  - 2009 IVOA – Massimo+Peppe – DAME, KDDIG Charter

CV and Experience

- 2010 DATAKON conference
- collaboration with Faculties of Informatics – consultant
    - 1 Bc, 1 unfinished PhD – FIT TU Brno
    - 1 Bc, 1 MS FI TU Ostrava (Java – SPLAT-VO)
- Official supervisor (1 unfinished PhD Masaryk University
    - astronomer with IT skills (CERN)
- 2011 FIT Czech Technical University Prague
    - External supervisor Bc and MS (~ 10)
- 2013-2015 Grant with informaticians (big data)
- 2014 COST Big-Sky-Earth
- 2015 External supervisor, 1 PhD on SW Engineering.
- + PhD, MS supervisor in astronomy (echelle pipeline)
- + oponent of theses at Masaryk Univ (NN , Chaos, ML)

CV and Experience

- Outreach - Virtual Observatory, AI
- Lectures + tutorials on VO (+AI) at universities (Serbia, Poland, Turkey, China, Czech Rep.)
- Acceptance of ideas and interest in VO+AI
  - Prof. astronomers – (very) little – afraid of IT
  - Prof. informaticians – they want to change all
  - Students of informatics– BIG !
  - Students of astronomy – little (no space for AI)
  - Public (open Doors, media … - Huge !)
  - Companies – in connection with EU grants - OK

Different Cultures

- Collaboration with (senior) informaticians
  - Where to publish (proceedings, conferences with peer-reviews x high ranked journal)
  - Don't care about astronomical interests – units, methods – just see their own field of interest in promotion of IT ideas
  - Publishing is paid (1000 USD/article) – WOS
  - They like books of Special Issues
  - They consider all astronomical data public
    - just to grab (no info about need of giving credits to telescopes, proprietary periods etc ...)
  - The ML is not science – formal grammar is !
  - They mostly do not have education in physics

Taxonomy of Students

- Students of Astronomy
  - Know programming of simple things
  - For work use C, F95, some Python
  - No SQL or RDBMS knowledge
  - No SW Engineering skills, no XML , no Java
  - Strong influence of their teachers (tools)
  - Little interest in instrumentation – give me data!
  - Very deep knowledge of Astrophysics – theory
  - Bad training in instrumentation – CCD, optics
  - During Bc and Master studies – study only
  - PhD. expected, stay at school, teach, externship

Taxonomy of Students

- Students of Informatics
  - Very good in SW Engineering
  - Formal information theory (compilers, grammar, parallelization, complexity - np-hard)
  - Use Python(Jupyter), cloud, java EE, C, Scala
  - Know SQL, RDBMS (+SPARK, Hadoop)
  - Basics of ML (some specialization even DL)
  - All work part-time – even at Bc. - for IT company
  - After graduation (even before) grabbed by head hunters , good positions in companies
  - PhD rare, only on theoretical issues, teaching ..

Considerations for Transmutation into Astroinformatician

- Student of astronomy
  - Must be technical geek in IT
  - Good programmer (Linux, Python, C) - hobby
  - He (not she) is little chaotic (many interests)
  - If he is good in IT, is also paid (part-time job)
  - Basics of statistics , chi^2, t, PCA, Bayesian ...
  - Some go to astronomy for hobby (they have already good jobs for years – age 35-50)
  - Most of young need scheduling in small steps
    - Don't have wider view of concept
    - Prefer hand work for small task (females more)
    - When finished with one task – consider all done

Considerations for Transmutation into Astroinformatician

- Student of Informatics
  - Advantage – interest in astronomy as young
  - Is able to learn physics on popularization level
  - Accustomed to have wide engineering view
  - No need to supervise them in IT (know best)
  - Very well self-organized (github, Jenkins..).
  - Little statistics (no advanced Bayesian)
  - Some have heard about R, not used it
  - If selects the AI – respects the science ethic
  - Is proud to have publication with astronomers

Probability of Successful Transmutation

- Astronomy has unique position in public
  - Its something fascinating to all people
  - Open doors experience – people know BH
  - Some residuals of ancient respect of nature
- Lot of students were interested in Astronomy
- Influence of amateur astronomical societies
- Strong correlation – astronomy + IT interest
- Great potential – University level education
- !!! will disapoint them
  - (a lot of theory x no or little observation)

Maximum Likelihood Justification

- Get them involved in astronomical problems for which no physics is needed (IT student) or no system engineering (Astronomy student)
- Explain them the problem in their language (spectrum=feature vector)
- Do not require the detailed documentation (be happy to have some)
- Let them work on IT stuff freely – they know best what they want
  - You do not know latest technology !
- Pay them from grant – small part-time jobs, exposure to institute spirit
  - Hard to keep them after graduation – IT is rich field
  - Maintain contact with them – invite them to convince novices
- Be their friend not supervisor
- Be prepared to adopt to their time constraints (work, lectures)
- Invite them to pub (beer, wine) – discuss topics there
- Let them collaborate on projects (common brainstorming, team meetings)
- They should know each other (personal enthusiasm of friend helps)

Maximum Likelihood Justification

- Motivate them for further collaboration
  - Put their thesis on ArXiv, Zenodo (slides, posters) DOI magic
- Expose the Student to IVOA
  - Nice community of astronomers/ IT geeks
  - Based on methodology/abstraction (no boring theory)
  - They meet the Big Science and see how it is done
  - Force him/her to presentation – Application WG best
- Expose them to ADASS (money restriction)
- Send them to (VO, AI) summer schools (PennState, India, Strasbourg)
- Big Sky Earth meetings – new opportunity

Realized Theses

- Cutout + normalization on SSAP server
- VO image server (SIAP + cutouts)
- VO catalogue + light curve server (700 mil measurements)
- Automatic catalog generation by parallelized clustering
- Automatic dynamic light curve generation
- IVOA Time series and data cube standard (ongoing)
- VO Cloud (server+workers for KOREL using UWS)
- Enhanced VO Cloud for ML
- Modules for RF, SOM
- Deep learning – modification of CAFFE on GPU
- Outlier analysis of milion of LAMOST spectra
- (Semi-)Supervised learning of Be star profiles
- Signed for – Deep Learning of spectra, VO-Cloud on SPARK, Jupyter
- SPLAT-VO enhancement, DM for time series, Jenkins

Conclusions

- It is easier to convert into AI the student of informatics
- The temporary success rate is high (2 years)
- The long term stability is low (IT jobs well paid)
- Results are satisfactory (in my case)
  - 1 enrolled in PhD at FI TU Prague, involved in IVOA
  - 1 PhD on motherhood leave … pending
  - 1 PhD pending – urgent tasks in IT...
  - 1 (SPLAT) in company interested in EU AI collaboration
  - 4 work in IT and are able to meet with newcomers
  - 1 running MS, 1 new Bc
  - 1 running MS – astronomer of my colleague
  - 3 stoped to continue in AI projects after Bc.

Conclusions

- The school must prepare new Data Scientists
- Definitely !

  To grow them into AI early before Bc.

  Push on Universities - change lectures
- Agreements of collaboration with Academy
- AI is appealing and challenging
- Interdisciplinarity - funding agencies
- Heavy task to start things to move

But WE MUST KEEP TRYING IT !

Astroinformatics is Fun !

Thank you !